Catalogue No. 97-09

SLID GEOGRAPHY AND ITS IMPACT ON LOW INCOME MEASUREMENT

Product Registration Number 75F0002M

August 1997

Ron Cunningham, Household Surveys Division

Pierre Lafrance, Household Surveys Division

John Rowland, Household Surveys Division

Joanne Murray, Household Surveys Division

The Income and Labour Dynamics Working Paper Series is intended to document detailed studies and important decisions for the Income and Labour Dynamics program. It is a continuation of the SLID Research Paper Series. These working papers are available in English and French. To obtain a summary description of available documents or to obtain a copy of any, please contact the Dissemination Unit at Jean Talon Building, 7th floor, section B5, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6, by INTERNET (DYNAMICS@STATCAN.CA), by telephone (613) 951-7355 or toll free 1-888-297-7355, or by fax (613) 951-3012.

ACKNOWLEDGEMENTS

The staff of SLID wish to thank Peter Bissett and Ingrid Ivanovs of Geography Division, and the staff of GEO-Help, for their assistance in the development of the methodology described in this paper.

EXECUTIVE SUMMARY

Geography is an important part of the Survey of Labour and Income Dynamics (SLID), from the geographical framework used by the sampling design to the geographic codes that facilitate the analysis and dissemination of the data. Geography helps to identify movers, and allows the comparison of socioeconomic characteristics between one region and another. It also enables analysts to link SLID data with other data sources.

The method of measuring low income most commonly used by Statistics Canada —the LICOs, or low income cutoffs—are dependent on geography for their derivation, and are sensitive to any variations in defining or applying the appropriate geographical concepts.

This working paper has three main objectives:

- to present the nature and uses of the geographic structure used by SLID
- to explain how statistics on low income are dependent on geographical concepts
- to describe the methodology for deriving geographic data for the Wave 1 (1993 reference year) SLID files, and improvements made for the Wave 2 (1994 reference year) release

106-02837/

TABLE OF CONTENTS

		Page
1.	PURPOSE OF PAPER	1
2.	THE GEOGRAPHIC COMPONENTS OF SLID	1
2.1	LFS frame geography	1
2.2	Information collected from movers	4
2.3	Postal codes: The "common denominator"	5
2.4	Data uses of geography information in SLID	6
3.	CODING GEOGRAPHY IN WAVE 1	9
3.1	Approach	9
3.2	Problems encountered	10
4.	NEW APPROACH	14
4.1	Objectives	14
4.2	How method was developed	15
4.3	High-level outline of method	17
5.	CONCLUSION	20
	DEFINITIONS	21

1. PURPOSE OF PAPER

The purpose of this working paper is to describe the nature and uses of the geographic structure used by the Survey of Labour and Income Dynamics (SLID). Aside from this general objective, the paper will focus on the question of how statistics on low income are dependent on geographical concepts.

The methodology for deriving geographic data for the Wave 1 (1993 reference year) SLID files, and improvements made for the Wave 2 (1994 reference year) release, are also explained.

2. THE GEOGRAPHIC COMPONENTS OF SLID

2.1 LFS frame geography

Before discussing what geographical information is collected in SLID, it is helpful to examine the geographical infrastructure used in the collection process itself.

The sample for each SLID panel is a subset of the Canadian Labour Force Survey (LFS) sample. The LFS sample is divided into six representative subsamples, called rotation groups. Once selected, a dwelling is retained in the sample for six months, and each month, one-sixth of the sample is replaced or "rotated". To obtain a subset of the LFS sample for SLID's Panel 1, two of the six LFS rotation groups, consisting of about 20,000 households, were selected. About 88% of LFS respondents, or 17,000 households, agreed to provide information for SLID. About 2,000 households were dropped for cost reasons, leaving a sample of approximately 15,000 households or 31,000 respondents aged 15 and over.

Sample selection for Panel 2 was conducted in a similar fashion by selecting two LFS rotation groups in January 1996.

The LFS sample is drawn from an area frame (see **Definitions**) and is based on a stratified, multi-stage design that uses probability sampling. The successive stratification and sampling stages are outlined below. Further information can be obtained in *Guide to Labour Force Survey Data*, Statistics Canada Catalogue No. 71-528-XPB and *Methodology of the Canadian Labour Force Survey, 1984-1990*, Statistics Canada Catalogue No. 71-526-XPB.

The LFS redesigns its sample every ten years to take advantage of the latest available census counts and characteristics. SLID's Panel 1 reflects the 1981 Census-based design in effect when the panel was selected. However, for Panel 2, the 1991 Census-based design applies. Apart from the increased efficiency of the new design, there are features that are likely to be beneficial from SLID's perspective. One is that total income was added as a stratification variable.

First, each province is divided into LFS economic regions (see **Definitions**), which are set by federal-provincial agreements. In the 1991 Census-based design, these are further subdivided by Employment Insurance Regions (EIRs), which are defined by Human Resources Development Canada (HRDC). Each of the resulting strata is in turn divided into one or more "urban" areas, one "rural" area, and various "special areas" (where applicable).[1]

Generally, the "urban" areas are large or medium sized urban centres. For the most part, they respect geographical boundaries established for the census, such as

[1] Note that the use of "urban" and "rural" in the LFS context is different from the standard census concepts as presented in the **Definitions** section.

census metropolitan areas or census agglomerations (CMAs/CAs - see **Definitions**). The area of the stratum not covered by these urban centres is considered "rural".

Each selected urban centre is further divided into strata, such that each stratum has a relatively homogeneous population in terms of socioeconomic characteristics as identified in the census. The number of strata depends on the size of the urban centre. Within each of the strata, groups of dwellings are identified, generally city blocks, groups of blocks or block-faces (see **Definitions**). These "clusters" are the primary sampling units (PSUs) for the urban centre. The selection of a sample of clusters from each of these strata represents the first stage of sampling in these areas.

In the largest urban centres, a sample of units in large apartment buildings is selected from a separate register based on information supplied by Canada Mortgage and Housing Corporation. The purpose of this is to ensure better representation of apartment dwellers in the sample, as well as to minimize the effect of growth in clusters due to construction of new apartment buildings.

For the 1981 Census-based design within the "rural" areas, further strata were delineated based on socioeconomic characteristics. These were the PSUs, serving the same purpose as city blocks in the urban centres, but covering relatively large geographic areas. Several of these areas were selected from each stratum in the first stage of sampling. Within each selected PSU, several clusters were then formed using well-defined physical features which were recognizable both on maps and in the field (rivers, roads, etc.). In the 1991 Census-based design, the extra stage of sampling was eliminated in most of the country, and the PSU is most often the enumeration area (EA - see **Definitions**).

In the 1981 Census-based design, the special areas were made up of the LFS target population found in structures excluded from the dwelling frame—for example, live-in staff of hospitals or schools, permanent residents of hotels or motels, and civilian personnel living on military bases—and some areas of the province that are too remote to be readily accessible to LFS interviewers. For the 1991 Census-based design, the special area frame consists solely of the remote areas—all the "collective dwellings" have been integrated into the regular design. EAs also serve as the primary sampling units for special areas.

In all three types of areas, selected clusters (or EAs) are visited by enumerators in the field, to prepare a listing of all private dwellings in the cluster. From the listing, a sample of dwellings is then selected for interviewing.

2.2 Information collected from movers

The movement of household members to a new location can be a deceptively complex event, since it could involve a move of the entire household or of just a part of it. If only some of the members move, they could be joining another existing household or creating a new one. If a new household is created, it may have been formed by members coming from the same household or from separate ones. As well, a move may be a gradual transition rather than an event, such as what can occur with a marital separation where a "trial separation" may lead to a permanent one. The following variables produced by SLID attempt to describe geographical mobility:

- number of times person moved in the reference year
- date of last move
- reason moved

- nature of last move, i.e. entire household move or household split
- flags indicating whether the move occurred within the same province, the same CMA/CA, or the same census subdivision (CSD - see **Definitions**)
- flags indicating whether the household composition has changed, the household or the address has changed, or a new household member is present
- flags indicating whether the family composition has changed, the family or the address has changed, or a new family member is present (see **Definitions** under "Economic family")

2.3 Postal codes: The "common denominator"

A respondent's residence is linked to geographic areas at varying levels. Though only province and region are available on the external file, the SLID database contains identifiers for urban area (see **Definitions**), CMA/CA, EA, and CSD, among others. These are not collected directly from the respondent but are "rolled up" from the basic geographic building block used in SLID processing, the postal code (see **Definitions**).

The postal code is a good choice as building block since virtually all respondents know theirs, and because it is usually small enough, at least in more urbanized areas, to determine unambiguously all other geographic areas of interest. Since 1983, Statistics Canada has maintained the Postal Code Conversion File, or PCCF (see **Definitions**), which contains the link between postal codes and Statistics Canada's census geographic codes.

In the majority of cases, the postal code corresponds to the location of the respondent's residence; however, in 1.7 % of cases across Canada, it refers to the location of a post office where the household has a box to receive their mail.

For other geographic limitations of the postal code, see the **Definitions** section.

The methodology for deriving Statistics Canada geographic codes using the PCCF is described later in the paper.

2.4 Data uses of geography information in SLID

The wealth of geographic data present on the internal SLID file permits a wide range of uses:

The population of a specific area or areas at a given point in time may be profiled in terms of their characteristics available on the SLID files. Due to the relatively small sample size used in SLID, there are significant limits involved in this type of analysis.

The EA is the basic geographic unit of the census. Because SLID records are linked to EAs through the postal code, a relationship can be made with the census database. For example, suppose a user wanted to analyze the relationship between earnings and knowledge of official languages. This particular language variable is not available in SLID[2], but it is collected by the census. Although it would not be possible to link census data and SLID data for each person, the user could obtain

---

[2] although *mother tongue* is. Mother tongue is defined as the first language learned at home in childhood and still understood by the individual.

the most recent census figures by EA and make a correspondence to the SLID files through the EA code.

In a similar way as for census data, variables available from other surveys, if tabulated by EA or by some aggregate of EAs, could be linked to data from SLID.

Because the SLID database contains the Employment Insurance Region code[3] (EIR - see **Definitions**), unemployment data available by EIR could be linked to SLID data. For example, persons who moved from one EIR to another with a different unemployment rate could be examined in terms of such characteristics as number of jobs held or amount of EI benefits collected.

The measurement of low-income incidence also depends on geography. The Survey of Family Expenditures (FAMEX) produces statistics on Canadian families' expenditures on food, shelter, and clothing. Based on these figures, an average level of these expenditures as a percentage of total family income is calculated. Twenty percentage points are added to this figure, the idea being that any family that spends this high a proportion of their income on the necessities of life could be considered to be "low income". Family size and the degree of urbanization of a family's residence are factored into the measurement and a set of low income cutoffs (LICOs) is produced. These represent the incomes below which a family is put into the low income category. The cutoffs are updated each year based on changes in the Consumer Price Index. Based on the 1992 FAMEX, the average family expenditure on food, shelter, and clothing was 34.7% of total (before-tax) income, which means the LICOs were based on an expenditure level of 54.7% (i.e.

---

[3] The initial Wave 1 release did not contain the EIR codes. However, during Wave 2 processing the codes were added for reference years 1992 through 1994.

34.7% + 20%). Separate sets of LICOs are produced for before-tax and after-tax income.

The degree of urbanization for deriving the LICOs is determined using the variable "size of area of residence" (see **Definitions**), which represents the population of the area in which the respondent lives. The *SLID Microdata User's Guide* for Wave 2 shows the LICOs for reference years 1993 and 1994. The SLID database contains the following variables relating to low income:

- low income cutoff (LICO) for the family (before taxes)
- flag indicating whether family is below the LICO (before taxes)
- ratio of family income to LICO (before taxes)
- low income cutoff (LICO) for the family (after taxes)
- flag indicating whether family is below the LICO (after taxes)
- ratio of family income to LICO (after taxes)

For more information on the LICOs and for statistics on incidence of low income, refer to *Income Distributions by Size in Canada*, Statistics Canada Catalogue No. 13-207-XPB (annual).[4]

Finally, the SLID files contain latitude-longitude coordinates for each record[5], based on the postal code of the respondent's residence. This permits the user to

---

[4] The 1990 edition of this publication also describes an alternative method for measuring low income, known as the low income measure (LIM), which is based on one-half the median family income, adjusted for family size. Because this measure is independent of geography, it is not discussed in this paper.

[5] As with the EI region codes, the Wave 1 release did not contain latitude/longitude, but it was added for reference years 1992 through 1994 during Wave 2 processing.

calculate the approximate distance moved by the respondent (if applicable), as a measure of geographical mobility. Provided in the SLID Microdata User's Guide for Wave 2 is a formula that permits calculation of the distance between two points given in latitude-longitude (the "great circle" distance).

3. CODING GEOGRAPHY IN WAVE 1

3.1 Approach

For Wave 1, all microdata records, with the postal codes that respondents had provided, were sent to Geography Division to be matched against the PCCF. In this way, 1991 Census geographic codes were added to all records.

In some cases, a postal code was linked to two or more EAs. This occurs where the area served by a postal code is spread over (parts of) several EAs, which happens particularly in rural Canada, where postal code areas are large compared to EAs. When this was the case, one of the linked EAs was selected randomly.

In other cases, the postal code was not present, was an invalid format (i.e. other than A9A 9A9), or was not present on the version of the PCCF for the reference year in question. Where no match was found for any of these reasons, Geography staff used whatever relevant information was available, such as street name, telephone number, or municipality, to assign a federal electoral district (FED - see **Definitions**) and EA code. In a few cases, no such locational information was present, and a FED/EA code was selected at random within the appropriate province.

3.2 Problems encountered

One of the major problems encountered was due to the random assignment of an EA within a postal code area, and its effect on measurement of low income. The Low Income Cutoff (LICO), described in Section 2, is a function of the population of the urban area ("size of area of residence") in which the respondent lives. Improper identification of the urban area, if it happens according to a particular pattern, can distort the calculation of low-income incidence. This would have occurred in the processing of Wave 1 data, as described below.

EAs in rural areas are, in general, less populated than those in urban areas. According to the 1996 Census, urban EAs had an average population of 708, while rural EAs averaged only 362 (Alternatively, EAs located within CMAs or CAs averaged a population of 689, versus 381 for other EAs.). Thus, given the random assignment of multiple EAs linked to a postal code, persons living in rural EAs would tend to be over-represented. This would have occurred in many of the cases where these multiple EAs were a mixture of urban and rural (Out of a total of 8,989 unique postal codes on the Wave 1 files, 1,486 were linked to multiple EAs. In 17% of these instances the multiple EAs were all urban, in 41% they were all rural, and 42% of the time there was a mixture of urban and rural.[6]). This situation would have caused a disproportionate number of respondents to be assigned to rural areas, and therefore to be given a lower ("more rural") LICO

[6] Tabulations from the Wave 1 files also show that in urban areas it is very uncommon for the multiple EAs to be located in different CMA/CAs (or, some inside a CMA/CA with some outside), while in rural areas it is fairly common. It is also much more common for rural EAs to be in different CSDs.

than they should have.[7] The result would therefore have been an overall underestimate of low income incidence.

Another problem was recognized which would have resulted from the random selection of an EA where there was a one-to-many relationship: since there was no guarantee that the same EA would be selected year after year for the same dwelling, there were cases where it appeared that the respondents had moved when they had not. That is, blind repetition of the process from year to year would have led to an artificially high level of geographic mobility.

SLID Research Paper No. 96-07 documents a discrepancy in estimates of low income between SLID and the Survey of Consumer Finances (SCF). Comparison of 1993 reference year data revealed a lower incidence (3 percentage points) of low income Canada-wide in the SLID data compared to SCF. While most of the discrepancy can be attributed to differing distributions of family income levels, CPP/QPP income, and estimates of family size, about one-quarter of it is probably due to a difference in the derivation of the size of area of residence within a CMA/CA. SLID Wave 1 processing used the census definition, which distinguishes between separate urban areas, as well as the residual rural area, within a CMA/CA. In SCF, all locations within the CMA/CA are assigned a size of area of residence corresponding to the total CMA/CA population.[8] The effect is

---

7 In the same vein, some sets of multiple EAs could have all been urban but belonged to distinct urban areas of differing populations, causing the record to be linked to the wrong size of area of residence and therefore given the wrong LICO. Again there would presumably be a rural or, at least, "less urban" bias occurring. However, for Wave 1 there was only one case where multiple EAs were in 2 distinct urban areas.

8 The derivation of size of area of residence differs in another way. In SLID, the postal code is used to link to the appropriate urban area, while in SCF, the Labour Force Survey sample unit identifier is used.

that, for SLID Wave 1 as compared to SCF, some respondents living in a CMA/CA would have a higher probability of being assigned a smaller size of area of residence, therefore a lower LICO, and therefore would have a lower probability of being put into the low income category.

Table 1 shows the distribution of population by size of area of residence, according to 1993 SCF and SLID Wave 1. The SLID data are adjusted to include "Don't know" in the other categories, and further adjusted to conform to the same size of area of residence as SCF within CMAs and CAs. After adjustment for SCF size of area of residence, the SLID population distribution conforms more closely to the SCF distribution. This is confirmed by comparing the weighted average percentage point deviations.[9] The SLID figures (unadjusted for SCF size of area of residence) generate a deviation of 3.0 compared with SCF. However, after the adjustment is made, the deviation is only 1.6. In other words, about half the difference in the population distributions is explained by the difference in concepts for size of area of residence (as opposed to other factors such as differences in sampling or collection procedures).

Another characteristic of the Wave 1 processing was that, for all non-response records, no imputation of any of the geographic code fields was done. Since it is not clear that non-response is entirely independent of geography, this situation may also be a cause of distortion in the data.

---

[9] The weighted average percentage point deviation is an index number that gives a measure of fit between two sets of percentages. It is obtained by calculating the absolute differences between the percentages in each row of the table, then computing a weighted average by weighting the average deviation in each row by the population living in those areas (according to SCF).

Table 1

Population by size of area of residence

(1993[1] Survey of Consumer Finances, 1993[1] Survey of Labour and Income Dynamics)

	SCF		SLIDa[2]		SLIDb[3]		Percentage point deviations (weighted by SCF figures)	
Size of area of residence	population (1000's)	%	population (1000's)	%	population (1000's)	%	SLIDa vs SCF	SLIDb vs SCF
Rural	5,045	17.7	6,535	23.0	5,196	18.3	26,943.99	3,117.50
0-29,999	3,607	12.6	3,537	12.5	2,713	9.6	656.66	11,138.28
30,000-99,999	2,645	9.3	2,776	9.8	2,531	8.9	1,346.27	943.01
100,000-499,999	4,159	14.6	3,537	12.5	5,263	18.5	8,805.70	16,470.16
500,000 +	13,068	45.8	11,994	42.3	12,683	44.7	46,396.70	14,813.56
Total	28,524	100	28,379	100	28,386	100	**3.0**	**1.6**

[1] 1993 **reference year**

[2] adjusted for "Don't Knows"

[3] adjusted for "Don't Knows" and size of area of residence

4. NEW APPROACH

4.1 Objectives

The problems and limitations described in the preceding sections demanded an improved process for succeeding waves. The development of this process would be guided by the following objectives relating to geography:

- Compatibility: The geography used by SLID should "plug into" established geographies, permitting comparability with data from other sources, e.g. census, SCF, and FAMEX.
- Repeatability: The procedure should yield the same results year to year when provided with the same inputs.
- Minimal intervention: The process should be as automatic as possible with minimal requirement for manual action or judgment.
- Completeness: Records with total or partial non-response, or invalid response, should have geographic codes at all levels. Wave 1 records with missing or invalid geographic data (for reference years 1992 and 1993) should be fixed retroactively.

In addition to these general principles, it was desired that SLID staff have more control over the geographic processing, rather than just passing the inputs to Geography Division and receiving back the results. This would increase the level of geographic knowledge within Household Surveys Division and permit more control over scheduling and decisions to be made along the way.

4.2 How method was developed

The development of a new method for deriving geographic codes for Wave 2 focussed on overcoming the limitations previously described, while adhering closely to the objectives stated above. To enhance the compatibility with other geographies, it was decided to use an "adjusted" size of area of residence variable based on the SCF methodology to generate low-income statistics. As described in Section 3, the size of area of residence for SCF is based on the population of the entire CMA/CA (if applicable) rather than just the urban area.

The desire for repeatability of results led to the investigation of how we could avoid the random selection from a list of several PCCF records linked to a postal code. On the PCCF is a flag called the "single postal code indicator", which points to the "best" record to use where more than one is linked to the same postal code.[10] This indicator represents an attempt to point out the record corresponding to the greatest concentration of population for the postal code area in question. It was decided to use this single postal code indicator in the process of assigning geographic codes, thus eliminating the random element and ensuring that respondents who had not moved would be assigned the same geography from one year to the next.

Section 3.2 described the downward bias of low income incidence occasioned by the size of area of residence concept, as well as by the random assignment of an EA for a multiple-record postal code. Using the original (Wave 1) method of assigning geography, the 1993 low income incidence, before tax, of persons 16 years and over, was 14.7%. Using the new method applied to the 1993 SLID

---

[10] Out of a total of 689,152 active postal codes in the most recent version of the PCCF, 31,681 (4.6 %) appear on multiple records.

data, the incidence increased to 15.9%, moving it closer to the SCF figure for that year (for persons 15 and over) of 18.0%.

The desire to reduce manual intervention (for purposes of both cost savings and repeatability) led to the consideration of other electronic data sources. One file that was considered was the 1991 Place Name Master File (PNMF), a listing of all "place names" (see **Definitions**) recorded in the census, including unincorporated places, alternate names, and historical places, in addition to legally incorporated municipalities. The place names are linked to a CSD, and in some cases to an EA. It was hoped that the PNMF would help in determining an EA code based on a place name where no postal code was present. However, because a significant number of the place names had no EA link, or links to multiple EAs, it was decided not to use this file in view of its limitations.

Telephone billing files (see **Definitions**), contain the telephone numbers and addresses, including postal codes, of telephone company subscribers, and thus serve as a good source for records where there is no postal code (Since SLID interviews are conducted by telephone, there is always a phone number associated with a respondent.). These files were actually used in Wave 1 by Geography Division, for some cases with missing or invalid postal codes, but for Wave 2 their use was made a standard step in the process executed within Household Surveys Division.

To meet the objective of completeness in the data, all non-response records would have an EA code imputed, which would allow the "roll-up" to all higher geographic codes as well. Thus, there would be no "Don't know" values for any of the geographic areas.

The key means of gaining more control over the process within Household Surveys Division was the creation of a SLID-specific Postal Code Conversion File, which would be a stripped-down version of Geography's PCCF, created by Household Surveys Division and designed specifically for the purpose of deriving geographic codes for SLID data. The creation and use of this file is detailed in the next section.

4.3 High-level outline of method

For reference year 1992, the following steps were done with the exception of the first part of Step I (the year-to-year matching). The entire process was then repeated for the 1993 and 1994 reference years. Also, match statistics for each step in the 1994 processing are indicated after the descriptions.

(I) The latest SLID household-based file is matched against the one from the previous year, by household ID, province of residence, and postal code. For all matched records, census geographic codes are copied over from the corresponding old records. This ensures that households with the same postal code in both years will have the same geographic codes.

Included in this match are households where the postal code was imputed the previous year. If there is a match on household ID and province, if the postal code for a household was imputed in the previous year, and if the postal code for the current year is blank or invalid, the imputed postal code, plus all the other geographic codes, are carried over to the current year. This ensures that the household is not randomly moved to a different geographic area, through re-imputation, when in fact no move took place.

For all records, the postal code and telephone number are checked for validity against the province; that is, the province is assumed to be correct, and if the postal code and/or telephone number do not match, they are blanked out.

Finally, for all unmatched records, the household ID, reference year, province, address, city, telephone number, postal code, and imputation flag are extracted and sent to step (III).

[11,085 records out of 15,006 records, or 74%, were matched in this step for Wave 2]

(II) A special-purpose SLID-specific Postal Code Conversion File is created. This is a file at the postal code level containing all relevant geographic codes. Inputs to production of this file are the Postal Code Conversion File, Geographic Attribute File (GAF - see **Definitions**), and Employment Insurance Region file. The single postal code indicator is used to select a single EA record where more than one EA is linked to a particular postal code.

(III) Unmatched records from (I) are matched against SLID-specific PCCF from (II), by postal code. For all matched records, all geographic codes are copied over. Unmatched records are passed on to (IV).

[3,284 records out of 3,921 records, or 84%, were matched in this step for Wave 2]

(IV) Unmatched records from (III) are matched against the telephone billing files by telephone number. For matched records, the postal code is added and the record is re-matched against the SLID-specific PCCF to add all geographic codes. Unmatched records are passed to Geography Division for resolution.

[197 records out of 637 records, or 31%, were matched in this step for Wave 2]

(V) All remaining unmatched records from (IV) are checked in Geography Division for a valid address or any other information (street name, description of location or dwelling, municipality) that would help to locate the dwelling as precisely as possible within the province (which is always known). Based on this information, an FED/EA code (see **Definitions**) is assigned or imputed. The records are then matched against the SLID-specific PCCF, by FED/EA, to obtain the remaining geographic codes.

[134 records out of 440 records, or 31%, were assigned FED/EA code based on locational information at the sub-provincial level.]
[289 records out of 440 records, or 66%, had FED/EA code imputed]
[17 records out of 440 records, or 4%, referred to residences outside Canada]

(VI) All matched/resolved records from the previous steps, with geographic codes added, are combined into a single file.

5. CONCLUSION

The results of SLID Wave 1 processing provided a stimulus for improvements to the method of assigning geography for the second wave. These improvements resulted in more harmony with other data sources, particularly with SCF, increased year-to-year consistency, decreased manual intervention, more completeness, and more control over the process by SLID staff.

The availability of the Wave 2 data will provide the first opportunity for longitudinal analysis of most SLID variables. This analysis may well bring to light new challenges, along with ideas on how to further improve the geography of SLID.

DEFINITIONS

Note: Since all SLID releases to date have been based on 1991 Census geography, all definitions of census geographic areas have been taken from the *1991 Census Dictionary* (Statistics Canada Catalogue No. 92-301).

Area frame

A division of the entire population under study into a number of distinct geographical areas. A sample of these areas is then selected, and a listing of population elements is obtained/created for each one (Usually, a second stage of sampling then takes place when a sample of the elements is taken for each area.).

Block-face

The general concept of a block-face is one of a small recognizable geographical unit to which census data can be associated. The goal is to approximate, through aggregation, user-defined query areas for census data extraction and tabulation.

The block-face refers to one side of a city street, normally between consecutive intersections with streets or other physical features (such as creeks or railways).

Census agglomeration (CA)

The general concept of a census agglomeration (CA) is one of a large **urban area**, together with adjacent urban and **rural areas** which have a high degree of economic and social integration with that urban area.

A CA is delineated around an urban area (called the urbanized core and having a population of at least 10,000, based on the previous census). Once a CA attains an urbanized core population of at least 100,000, based on the previous census, it becomes a **census metropolitan area (CMA)**.

Census division (CD)

Refers to the general term applying to geographic areas established by provincial law, which are intermediate geographic areas between the **census subdivision** and the province (e.g., divisions, counties, regional districts, regional municipalities and seven other types of geographic areas made up of groups of census subdivisions).

In Newfoundland, Manitoba, Saskatchewan and Alberta, provincial law does not provide for these administrative geographic areas. Therefore, census divisions have been created by Statistics Canada in co-operation with these provinces.

Census metropolitan area (CMA)

The general concept of a census metropolitan area (CMA) is one of a very large **urban area**, together with adjacent urban and **rural areas** which have a high degree of economic and social integration with that urban area.

A CMA is delineated around an urban area (called the urbanized core and having a population of at least 100,000, based on the previous census). Once an area becomes a CMA, it is retained in the program even if its population subsequently declines.

Smaller urban areas, centred on urbanized cores of a population of at least 10,000, are included in the **census agglomeration (CA)** program.

Census subdivision (CSD)

Refers to the general term applying to municipalities (as determined by provincial legislation) or their equivalent, e.g., Indian reserves, Indian settlements and unorganized territories.

In Newfoundland, Nova Scotia and British Columbia, the term also describes geographic areas that have been created by Statistics Canada in co-operation with the provinces as equivalents for municipalities.

CMA/CA parts

The concept of CMA/CA parts distinguishes between central and peripheral **urban** and **rural areas** within a **census metropolitan area (CMA)** or a **census agglomeration (CA)**. There are three CMA/CA parts: urbanized core, urban fringe and rural fringe.

Urbanized core: A large urban area around which a CMA or a CA is delineated. The urbanized core must have a population (based on the previous census) of at least 100,000 in the case of a CMA, or between 10,000 and 99,999 in the case of a CA.

Urban fringe: An urban area within a CMA or CA, but outside the urbanized core.

Rural fringe: All territory within a CMA or CA lying outside urban areas.

Economic family

Refers to a group of two or more persons who live in the same dwelling and are related to each other by blood, marriage, common-law or adoption.

Economic region

Labour Force Survey (LFS) economic regions are subprovincial areas of similar economic structure, used as the primary stratum for the Labour Force Survey. Identical to **subprovincial regions** except in British Columbia.

Employment Insurance Region (EIR) file

A file linking 1991 **enumeration areas** to employment insurance regions (EIRs). Human Resources Development Canada uses EIRs in the administration of the EI program. Eligibility for EI is determined based on the applicant's employment record during the past year but also on the unemployment rate in the applicant's EIR.

Enumeration area (EA)

An enumeration area (EA) is the geographic area canvassed by one census representative.

EAs are to be as compact as possible to minimize travel and optimize census representative work. EAs are delineated so that the census representative may

locate them with as little difficulty as possible. Therefore, wherever possible, EAs follow easily recognizable physical features (such as the road network and rivers).

Enumeration area (EA) criteria include:

(1) Dwellings - the number of dwellings in an enumeration area generally varies between a maximum of 375 dwellings in large urban areas to a minimum of 125 in rural areas.

(2) Limits - an enumeration area never cuts across any geographic area recognized by the census.

Enumeration areas (EAs) are primarily census collection units; they are not designed as dissemination areas. Nonetheless, the EA is the smallest geographic unit for which census data are usually available.

EAs may change limits from census to census. Approximately 40% of the 1991 EAs remained identical to 1986 limits; however, the balance changed due to:

(a) population growth;
(b) a redistribution of **federal electoral district** (FED) boundaries in 1987;
(c) changes to geographic limits recognized by the census;
(d) the extension of the "mail-back" collection methodology into the 1986 "pick-up" areas;
(e) changes in delineation criteria.

Federal electoral district (FED)

A federal electoral district refers to any place or territorial area entitled to return a member to serve in the House of Commons (source: Canada Elections Act, 1990). There are 295 FEDs in Canada according to the 1987 Representation Order.

Federal electoral districts are defined according to the following criteria:

(1) the legal limits and descriptions are the responsibility of the Chief Electoral Officer and are published in the Canada Gazette;

(2) FED limits are usually revised every 10 years after the results of the decennial census.

1991 Geographic Attribute File (GAF)

The Geographic Attribute File assigns to each 1991 Census **enumeration area** all higher geographic levels by codes and names. It also locates each enumeration area representative point according to Universal Transverse Mercator (UTM) coordinates and Lambert coordinates, as well as by longitude and latitude. The 1991 population and private occupied dwelling counts for all enumeration areas are also presented.

For the purpose of assigning geographic data in SLID, the GAF is the source of data concerning **urban** and **rural areas**, specifically, the urban/rural flag, urban area code, and urban area population.

Place name

Place name is a general term for localities, urban neighbourhoods, post offices, communities and other types of unincorporated places (UPs). Place names include the names of inhabited places, formerly inhabited places, and other names associated with some human activity.

Postal code

(Note that the use of "urban" and "rural" in this definition is different from the standard census concepts as presented elsewhere in this section.)

The postal code is a six-character alpha-numeric code defined and maintained by Canada Post Corporation for the processing (sortation and delivery) of mail.

The alpha-numeric characters are arranged in the form ANA NAN, where "A" represents a letter of the alphabet and "N" a numeric digit. The first character of a postal code (allocated in alphabetic sequence from east to west across Canada) represents a province or territory, or a major sector entirely within a province.

The first three characters represent a set of well defined and stable areas known as the Forward Sortation Area (FSA). Rural FSAs are identifiable by the presence of a "0" in the second position of the FSA code.

The last three characters identify the Local Delivery Unit (LDU). In established urban areas, the LDU can specify a small and easily defined area within an FSA such as a **block-face** (one side of a city street between consecutive intersections with streets or similar physical features), an apartment building, an office building,

or a large firm or organization which does large volume business with the post office.

In rural areas, the LDU denotes a service area - the area serviced by rural route delivery from a post office or postal station, e.g., a rural route, general delivery or post office box.

The postal code represents a spatial referencing system which allows large volumes of geographic data to be manipulated, retrieved and analyzed at the micro-level.

When used as a linking tool between administrative files, the postal code protects the confidentiality of names or specific addresses, while at the same time offering a finer level of data aggregation.

Postal code (FSA) boundaries were not designed to respect standard geographic boundaries (except at the provincial level). Even then, there are three cases where the FSA valid in one province will service a **census subdivision (CSD)** within an adjacent province. This occurs when a CSD straddles a province boundary. Manitoba FSA "R8A" services both the Manitoba and Saskatchewan portions of Flin Flon. Saskatchewan FSAs "S9V" and "S0A" service the Saskatchewan and Alberta portions of Lloydminster. The Alberta portion of Makaoo 120 Indian Reserve is serviced by the Saskatchewan rural postal code "S0M 2E0".

Rural postal codes refer to the post office location. The actual service area of a rural post office is defined by the extent of its rural routes. Therefore, the inferred link to standard geography is according to the physical location of the rural post office and not the service area.

It is difficult to identify where a person lives based on a rural postal code. In some northern delivery cases, a rural postal code valid for the southern part of the province is also used to service northern settlements.

Urban postal codes representing a post office box also cannot be used to geo-reference a street location because the postal code is within a post office.

The community mail box has introduced an expanded territory for the postal code. In new growth areas, a community mail box postal code can now represent both odd and even sides of a street and different streets within a 300-metre radius of the community mail box.

Postal Code Conversion File (PCCF)

The Postal Code Conversion File (PCCF) links the six-character postal code with the standard 1991 Census geographic areas (such as **enumeration areas,** municipalities, census tracts, etc.). It also provides the x,y coordinates for a point representing the approximate location of the postal code to support mapping applications.

Rural area

The general concept of a rural area is that of a sparsely populated area.

Statistics Canada defines rural areas as those areas of Canada lying outside **urban areas.**

In addition to the above, within the area of Canada defined as rural, population densities and living conditions can vary greatly. Included in rural areas are:

- small towns, villages and other populated places under 1,000 population according to the previous census;

- rural fringe areas of **census metropolitan areas** and **census agglomerations** which may contain estate lots and other non-farm land uses, as well as intensive agricultural land uses;

- agricultural areas;

- remote and wilderness areas.

Size of area of residence

Refers to the classification used in tabulations where **urban** or **rural areas** are distributed according to the following size groups, based on their 1991 population:

Rural
Urban: Under 30,000
Urban: 30,000 - 99,999
Urban: 100,000 - 499,999
Urban: 500,000 - 9,999,999

Subprovincial region (SPR)

Refers to a subprovincial geographical unit smaller than a province (with the exception of Prince Edward Island and the Territories) made up of groupings of complete **census divisions.**

The subprovincial regions were created in response to the requirement for a geographical unit suitable for the analysis of regional economic activity. Such a unit is small enough to permit regional analysis, yet large enough to include a sufficient number of respondents, such that, after confidential data are suppressed, a broad range of statistics can still be released. The regions are based upon work by Camu, Weeks and Sametz in the 1950s with minor adjustments over the years to accommodate changes in census division boundaries and the views of provincial officials.

Subprovincial regions may be economic, administrative or planning regions. In some provinces, these regions are designated by law. In other provinces, the regions were created by agreement between Statistics Canada and the province or territories concerned.

The number of subprovincial regions varies extensively between provinces. For example, in Ontario there are five SPRs, compared to 16 SPRs in the province of Quebec.

The entire province of Prince Edward Island, the Yukon and the Northwest Territories consist of one subprovincial region each.

Subprovincial regions are used mainly in the dissemination of statistics on building permits, income tax data, manufacturing industries and employment data. The census respects these regions in order to provide data for comparative geographies.

Remarks: There are 68 subprovincial regions in all of Canada.

The 1991 SPRs coincide with the Labour Force Survey (LFS) **economic regions**, except in British Columbia.

For British Columbia, SPRs respect British Columbia's development regions. These regions are defined by B.C.'s Ministry of Regional and Economic Development to promote economic growth and to ensure access to government services in all regions of the province.

In Quebec, the 1991 SPRs respect Quebec's "régions administratives" defined by Quebec to serve as a basis for the production of statistics by government departments as well as the establishment of regional and local offices.

In all the other provinces, they have been delineated in co-operation with the provinces and respect LFS regions.

Telephone billing files

Updated telephone billing files are received on a quarterly basis by Statistics Canada from the various companies responsible for providing telephone service

across Canada. These files contain the addresses and telephone numbers of their subscribers, and provide a useful source of locational information for SLID respondents who have not provided a valid postal code. However, some limitations of the files for geographical purposes should be noted:

- The address reflects the mailing address where the subscriber's bill is sent, and thus will not always reflect the location of the person's residence. For example, it may be simply a post office box, which will have a different postal code from the person's residence.

- The phone number may relate to a person's temporary residence (e.g. for students or cottagers), while the address may be that of a permanent residence where the bill is sent.

- While business and residential numbers are usually distinguished from each other, this is not always true in the case of small businesses. Thus, the number may relate to the business location while the address reflects the location of the owner's home.

- Depending on the company, there may be no information provided for unlisted numbers, or such numbers may be scrambled, making these records unmatchable.

Urban area (UA)

The general concept of an urban area (UA) is that of an area containing a dense concentration of population.

Statistics Canada defines an urban area as an area which has attained a population concentration of at least 1,000, and a population density of at least 400 per square kilometre, at the previous census. All territory lying outside urban areas is considered **rural**. Taken together, urban and rural areas cover all of Canada.

Urban areas separated by gaps of less than two kilometres are combined to form a single urban area.